

Towards Computationally Creating Multi-answer Queries for the Remote Associates Test

Ana-Maria Olteteanu¹ and Kunkanit Yoopoo²

¹ Cognitive Systems, Bremen Spatial Cognition Center, Universität Bremen,
Germany

² Faculty of Information and Communication Technology, Mahidol University,
Thailand

Abstract. The Remote Associates Test is a creativity test used to assess human participants' ability for association. Small normative datasets of queries exist for this test; however, such datasets do not deal with the issue of potential multiple answers to the same test query. In this work we create a large dataset of queries to which multiple answers are possible. The computational work to create such a dataset is presented, together with the metrics relating to this dataset. The applications of this tool for the investigation and modeling of the creative processes of association in human cognition are also discussed.

1 Introduction

Imagine that, as a cognitive psychologist, you would want to investigate an aspect of the creativity and creative problem solving process in humans. Or that you would attempt to computationally model such a process. Various forms of tests exist to measure creativity and creative problem solving performance in human participants [4, 6, 3, 5]. However, some of these tests are old and do not provide normative data. Furthermore, such tests do not provide an ability to control for and parametrize their variables. Much more insight in the creative process could be obtained if cognitive psychologists and computational modelers would have access to large datasets of test items, the variables of which they could control. Furthermore, despite of measuring creativity, some such tests only allow for one "correct" answer, ignoring the fact that multiple answers might be possible, and thus being unable to explore how the cognitive process functions in the context of multiple solutions.

This paper starts from the premise that the investigation and testing of creative performance can benefit from the help of computational methods in establishing (i) new ways of assessing creative problem solving; (ii) better controlled parametrized stimuli sets for existing creativity and creative problem solving tasks and (iii) allowing and accounting for multiple possible solutions. The current work focuses on the last two: using computational methods to establish a set of controlled parametrized stimuli sets for a classical creativity test – the Remote Associates Test [7]. Specifically, we focus on computationally building and

extracting a set of Remote Associates Test queries for which multiple answers are possible.

The rest of the paper is organized as follows: the Remote Associates Test is briefly described in section 2, together with previous work on a computational solver for this test. An approach in creating stimuli subsets for multi-answer queries is described in section 3. The obtained dataset and the multi-answer query metrics are described in section 4. In closing, the applications of this work in cognitive psychology are discussed and future work is proposed.

2 The Remote Associates Test, comRAT-C and comRAT-G

The Remote Associates Test [7] (RAT) is a creativity test often used in the literature [2, 1]. In this test, three word queries are given to participants, like the query COTTAGE, SWISS, CAKE. The participants are asked to come up with a fourth word, which is connected to each of the query words. A potential answer in this case would be CHEESE. According to its creators, the RAT aims to measure creativity as the ability to make associations.

In previous work, [9] implemented a computational solver of the RAT called comRAT-C. This solver used language data (bigrams) from the Corpus of Contemporary American English, and a type of knowledge organization which supports the solving process [8, 11]. The solving of a RAT query can be visually represented as depicted in Figure 1. The initially given query words trigger word associates that have been previously encountered in conjunction with the query words. The query words shown in green in Figure 1 trigger the associates shown in blue. For example, the word CAKE triggers the words FLOUR and LAYER, because the cognitive agent has previously encountered expressions like CAKE FLOUR and LAYER CAKE.

Amongst the associates that are activated by each of the query words, some overlaps might happen. For example CHOCOLATE is such an overlap, triggered by associates of the query words SWISS and CAKE. The activation process started by presenting the query words will converge on such overlaps. A convergence of the associates of all three initial query words can result in an answer – like for example CHEESE in Fig. 1.

Besides solving the RAT computationally and correlating with human performance data, comRAT-C [9] has shown that multiple possible answers may exist for RAT queries, by sometimes providing different answers than the unique answer considered “correct” in the normative data. For example, for the query CHANGE, CIRCUIT, CAKE, the answer considered correct in the normative data was SHORT, while comRAT-C provided the equally plausible answer DESIGN. For the query HIGH, DISTRICT and HOUSE, the answer considered correct in normative data was SCHOOL, but comRAT-C provided other answers as well, like STATE, COURT, etc.

However, no dataset of queries with multiple answers was yet available. A researcher administering the RAT thus has no access to knowing whether her

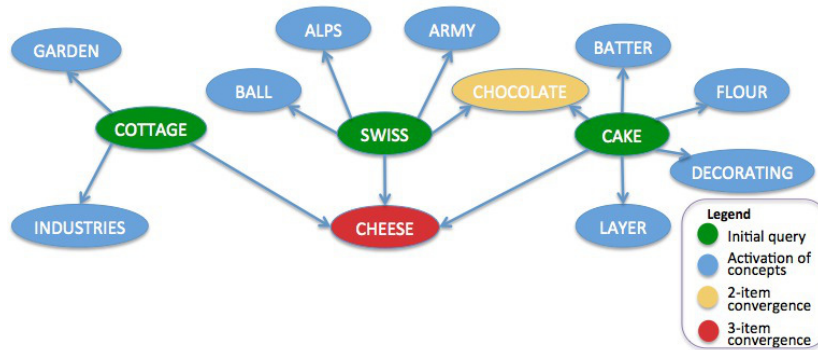


Fig. 1. A visual depiction of the associative process used by comRAT-C to solve the Remote Associates Test, at the concept level. Only a small subset of associates are depicted, in order to maintain visibility.

queries might potentially have different correct answers than the ones she is expecting. She might thus judge an answer as “wrong” just because this is not the answer expected as correct by the normative data. Meanwhile, this answer might be not wrong, but plausible, and different from the recognized correct answer. In comRAT-C computational terms, the participant might have just found a different convergence term, because of their knowledge base being structured or weighted slightly differently than that of other participants. As no account of multiple answers exists in the literature, however, this participant might end up with lower creativity scores because her answers do not match the “correct” answers, and this would affect the results of the empirical investigation.

Such plausible but different answers could also be used to investigate the process of solving this task at a deeper level. For example, why would one answer be preferred by a participant over another potential answer? Is this a function of that particular participant’s set of associations strengths in their memory/knowledge base? Or would certain associations be generally preferred over others? How would the parameters of such associations need to be modified in order to change the preferred answer? Manipulating various setups of queries with multiple answers could shed more light into the process of remote association. However, no hypothesis testing for queries with multiple answers is possible until a dataset for such queries is created.

3 Creating a Set of Multi-answer Queries

A set of 17 million RAT queries was created by reverse engineering the comRAT-C solving process with comRAT-G [10]. In short, this system considers each word as a potential answer, and uses its knowledge and knowledge organization to combinatorially generate queries which converge in that word as an answer.

Though very rich, this dataset is too large to explore manually, and requires the application of computational methods for extracting valuable subsets and

their metrics. In this work, we focus on the RAT queries which allow for multiple answers, and apply computational methods for finding all the multi-answer query sets, cleaning up this data computationally and building a multi-answer query dataset. We extract metrics regarding this dataset, as to prepare it for evaluation with human participants and distribution to the research community.

First, all multiple answer subsets are extracted. This step involved searching for query subsets of the form

$$\begin{aligned} &(w_1, w_2, w_3, ans_1), \\ &(w_1, w_2, w_3, ans_2), \\ &\dots \\ &(w_1, w_2, w_3, ans_x), \end{aligned}$$

where $w_k, k \in \{1, 2, 3\}$, stand for the various query words, and ans_x for the various potential answers. As shown in table 1, the application of this step has as result ordered subsets of queries which have multiple answers. For example, query ACCESS, BACK, SIDE is shown with both its answers PANEL and ROAD.

To offer the possibility of parametrising queries, the dataset we build also provides the following information for each query:

- the frequency of each of the query words – $fr(w_1), fr(w_2), fr(w_3)$;
- the frequency of the answer word, which might help differentiate between different answers to the same query – $fr(w_{ans})$;
- the frequency of the query words and answer words together as an expression $fr(w_1, w_{ans}), fr(w_2, w_{ans}), fr(w_3, w_{ans})$;
- the conditional probability for achieving each of the answers, given the query words ($P[w_{ans}|w_1]$), ($P[w_{ans}|w_2]$) and ($P[w_{ans}|w_3]$);
- the probability of finding a particular answer if all query words are equally weighted.

All parameters are calculated based on the frequencies provided with the Corpus of Contemporary American English bigram dataset.

In the second step, we build a dataset in which each query with multiple answers is uniquely represented, together with the number of answers we found for that query, and the following metrics:

- (i) lowest, highest and mean conditional probability of the different answers to the query, if each of the query words equally influenced the answer;
- (ii) lowest, highest and mean conditional probability given each of the query words, across the different answers and
- (iii) lowest, highest and mean frequency of the query words.

The dataset and metrics thus constructed look as depicted in Table 2. These metrics are provided in order to help cognitive psychologists or other users decide which query subsets to use, and thus tailor the subset to their research question or problem.

Table 1. Multi-answer query subsets, example data extract. The [...] symbol stands for columns in the table which describe parameters and have not been shown here because of table size constraints.

w_1	w_2	w_3	$answer$	[...]
Access	Back	Side	Panel	
Access	Back	Side	Road	
Industry	Management	Tax	Consultant	
Industry	Management	Tax	Estate	
Industry	Management	Tax	Expert	
Industry	Management	Tax	Hotel	
Industry	Management	Tax	Officials	
Industry	Management	Tax	Waste	

Table 2. Data and metrics on each query subset. Please note that at least four decimals are provided in the dataset, but these, together with other columns, were compressed here for the sake of visual depiction

Query	No. of answers	P(all) Low	P(all) High	P(all) Mean	P(w_x) Low	P(w_x) High	P(w_x) Mean	F(w_x) Low	F(w_x) High	F(w_x) Mean
Ability Education Skills	2	0.003	0.0345	0.0324	0.0016	0.0709	0.00324	41	639	226.333
Graduate University Programs	4	0.0045	0.0019	0.011	0.0013	0.0295	0.011	31	355	118.417
Youth Team World	3	0.025	0.028	0.027	0.0009	0.0711	0.02694	24	241	104.778
Business Company Management	9	0.0019	0.009	0.0042	0.0006	0.0245	0.0042	24	563	109.556

4 Results - metrics of the dataset

A dataset of 1206622 queries with multiple answers was obtained in step one. Out of these, 403341 queries were unique, as observed after agglomerating the data in step 2. The mean number of answers for the entire dataset was 2.27 ($SD = 0.77$).

Most of the queries obtained were two answer queries (332974), while a few sets of queries had between 17-30 answers (6 queries). The metrics pertaining to the number of queries are shown split in nine bands based on their number of answers in Table 3.

5 Discussion and Future work

This paper briefly presented our initial efforts in computationally constructing a set of queries with multiple answers for the Remote Associates Test.

One of the challenges of creating this dataset related to the presence of plurals in multiple query answers. Our task was to search for subsets of the form $(w_1, w_2, w_3, ans_1), (w_1, w_2, w_3, ans_2), [...], (w_1, w_2, w_3, ans_x)$. However, subsets of queries with two answers were encountered where the two queries and answers were of the form $(w_1, w_2, w_3, ans_1), (w_1, w_2, w_3, pl(ans_1))$, where $pl(ans_1)$ is the plural of the other answer. For example, we encountered the query DRAFT, MEMBERSHIP, PUNCH with both answers CARD and CARDS. We used a set of

Table 3. Dataset metrics based on number of answers

No. of answers	Number of queries
2	332974
3-4	61259
5-6	7045
7-8	1461
9-10	401
11-12	132
13-14	44
15-16	19
17-30	6

plural rules for English to find such queries. We then compressed plural and singular forms of queries in one data item, maintaining the singular form and calculating the mean of the probability and frequency metrics.

As we have now created a dataset of multiple answer queries, the next steps are as follows:

- to evaluate the dataset with human participants;
- to create a set of normative data – expressing accuracy, answer times and preferred answers for a subset of multi-answer queries;
- to use the dataset (and support the use of the dataset) in various cognitive science applications.

The dataset can be evaluated with human participants by checking (i) whether participants consider multiple answers to be indeed viable answers and (ii) whether empirical relations hold between the propensity of people to choose a particular answer and the probability, frequency or other factors associated with the various answers. As part of future work we also intend to show participants multiple possible answers and have them choose the one they find to be the most “appropriate”, in conditions in which the answer choices are similar or different in probability/frequency or other factors. This will help us investigate whether such factors have an impact in perceived appropriateness of answers, and whether similarity or difference in a particular factor influences the difficulty of the choice, affecting response times.

The creation of a normative dataset for multi-answer queries requires gathering data from human participants regarding response times, and the number of times the various answers are given. Whether human answers in the case of such queries cover all the potential multiple answers, or a very small subset of them, and for which queries and answers this manifests is also an interesting future empirical question.

Various applications of the use of such a dataset exist for cognitive psychologists. This tool and dataset can be used to design experiments that can capture which answers are preferred in various multi-answer conditions – for example

in cases in which the frequency, probability, beginning letter, or other parameters are varied. This dataset can thus be used as a means to establish and falsify various theoretical hypotheses about the creative process and the process of association.

After evaluating this dataset with human participants, we intend to provide it for scientific use via an online interface.

Acknowledgements

Ana-Maria Olteteanu acknowledges the support of the German Research Foundation (DFG) for the Creative Cognitive Systems Project OL 518/1-1 (CreaCogs).

References

1. Bourke, P., Shaw, H.: Spontaneous lucid dreaming frequency and waking insight. *Dreaming* 24(2), 152 (2014)
2. Cunningham, J.B., MacGregor, J.N., Gibb, J., Haar, J.: Categories of insight and their correlates: An exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *The Journal of Creative Behavior* 43(4), 262–280 (2009)
3. Duncker, K.: On problem solving. *Psychological Monographs* 58(5, Whole No.270) (1945)
4. Guilford, J.P.: *The nature of human intelligence*. McGraw-Hill, New York (1967)
5. Kim, K.H.: Can we trust creativity tests? a review of the torrance tests of creative thinking (ttct). *Creativity research journal* 18(1), 3–14 (2006)
6. Maier, N.R.: Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology* 12(2), 181 (1931)
7. Mednick, S.A., Mednick, M.: *Remote associates test: Examiner’s manual*. Houghton Mifflin (1971)
8. Olteteanu, A.M.: *Publications of the Institute of Cognitive Science*, vol. 01-2014, chap. Two general classes in creative problem-solving? An account based on the cognitive processes involved in the problem structure - representation structure relationship. Institute of Cognitive Science, Osnabrück (2014)
9. Olteteanu, A.M., Falomir, Z.: comRAT-C: A computational compound remote associate test solver based on language data and its comparison to human performance. *Pattern Recognition Letters* 67, 81–90 (2015)
10. Olteteanu, A.M., Schultheis, H., Dyer, J.B.: Constructing a repository of compound Remote Associates Test items in American English with comRAT-G. *Behavior Research Methods, Instruments, & Computers* (accepted)
11. Olteteanu, A.M.: From simple machines to Eureka in four not-so-easy steps. Towards creative visuospatial intelligence. In: Müller, V. (ed.) *Fundamental Issues of Artificial Intelligence*, Synthese Library, vol. 376, pp. 159–180. Springer (2016)